

# BEST AVAILABLE COPY PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-005975

(43)Date of publication of application : 12.01.2001

(51)Int.Cl.

G06T 7/20

G06F 3/00

H04N 5/00

H04Q 9/00

(21)Application number : 11-173747

(71)Applicant : VICTOR CO OF JAPAN LTD

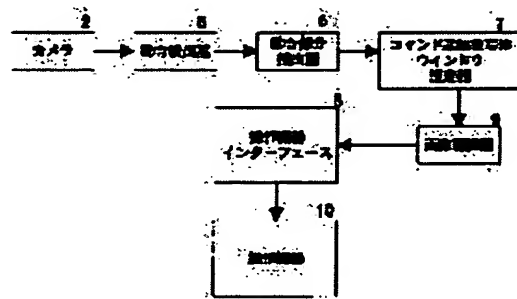
(22)Date of filing : 21.06.1999

(72)Inventor : SUGAWARA TAKAYUKI

**(54) DEVICE AND METHOD FOR CONTROLLING EQUIPMENT****(57)Abstract:**

**PROBLEM TO BE SOLVED:** To allow a user to control equipment without using an input device and a remote controller device.

**SOLUTION:** This device consists of a photographing means 2, a movement detecting means 5 which detects the movement of an image photographed by the means 2, a part extracting means 6 which extracts a moving part from a detected image, an image recognizing means 9 which recognizes the movement and/or shape of the image of the extracted part and an operation device interface means 8 which becomes an interface with an operation device 10 being an object. Then, the operation device of a control object is controlled by the recognized change of movement and/or shape.

**LEGAL STATUS**

[Date of request for examination] 28.09.2001

[Date of sending the examiner's decision of rejection] 15.03.2005

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection] 2005-006671

[Date of requesting appeal against examiner's decision of rejection] 14.04.2005

[Date of extinction of right]

THIS PAGE BLANK (USPTO,

THIS PAGE BLANK (USPTO,

\* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

CLAIMS

---

[Claim(s)]

[Claim 1] A photography means and a motion detection means to detect a motion of the image photoed by said photography means, A motion partial extract means to extract a part with said detected image lost motion, An image recognition means to recognize said motion and/or configuration of the image of a part which were extracted, Appliance control equipment which consists of an actuation device interface means used as an interface with the target actuation device, and is characterized by said thing [ having been recognized ] which move and/or controls the actuation device of a controlled system by change of a configuration.

[Claim 2] A photography means and a motion detection means to detect a motion of the image photoed by said photography means, A motion partial extract means to extract a part with said detected image lost motion, A window setting means to set a window as the field of a predetermined motion part from said image, Appliance control equipment characterized by consisting of an image recognition means to recognize change of the image which exists in said window, and an actuation device interface means used as an interface with the target actuation device, and controlling the actuation device of a controlled system by change of the image in said window.

[Claim 3] The appliance control approach characterized by photoing a photographic subject, detecting a motion of said photoed photographic subject image, extracting a part with said detected photographic subject image lost motion, and controlling the device of a controlled system according to a motion of the image of this part, and/or change of a configuration.

[Claim 4] Appliance control equipment according to claim 1 or 2 characterized by using the image which judges the distance of the image which is moving and exists in a predetermined range from two or more images photoed by said two or more photography means.

[Claim 5] It is appliance control equipment according to claim 4 characterized by performing the judgment of said distance which is moving in said two or more images using an epipolar constraint.

---

[Translation done.]

**THIS PAGE BLANK (USPTO)**

\* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

DETAILED DESCRIPTION

---

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the appliance control equipment and the approach of controlling a device using the image information which detected the device especially with the camera about the equipment and the approach for carrying out remote control, motion information, or both information.

[0002]

[Description of the Prior Art] Conventionally, there is remote control using the infrared radiation used for television, video, etc. Moreover, when a CCD camera etc. is used, the motion of an image is recognized and a motion is detected, equipments, such as monitoring system which makes the image transcription of VTR linked to a camera start, exist.

[0003] As shown in drawing 16 , remote control indicated by JP,5-307793,A serves as an interface with a user, and consists of display / input section 100 which can change the contents of a display, a microcomputer (data-processing section) 200, the controlled-system sections, 400, such as VTR, and storage means 300, such as memory. And data processing of the instruction directed by display / input section 100 is carried out by the data-processing section 200, after interpreting an instruction, it performs, and the information for new display / input section 100 is saved in memory 300, and it can be made to carry out a busy condition using the information adjustable [ of the display / input section 100 ]. While using the controlled-system section 400, this learns a user's busy conditions (a peculiarity, the description, inclination, etc.), and it is invented in order to set it as the interface which is easier to use automatically.

[0004]

[Problem(s) to be Solved by the Invention] However, the carbon button and touch panel for an interface with a user were required of the above-mentioned conventional appliance control equipment as an input device. In such an input unit, the user needed to move in front of the input unit, and needed to operate it. Moreover, although the remote control (remote control) equipment which transmits an infrared control signal was required in order that a user might operate it, without moving in front of an input unit, there was no remote control equipment near the user, and the inconvenience on actuation was sometimes plentifully sensed by losing temporarily.

[0005] Furthermore, although there was the approach of controlling to detect a motion of the image photoed with the camera in monitoring system, and to start the image transcription of VTR, it was what is hard to be referred to as that this detects the existence of a motion of an image to the last, performs easy control, and has played a role of a user interface.

[0006]

[Means for Solving the Problem] A motion detection means 5 to detect a motion of the image photoed by the photography means 2 and said photography means 2 in order to solve the technical problem mentioned above, A motion partial extract means 6 to extract a part with said detected image lost motion, An image recognition means 9 to recognize said motion and/or configuration of the image of a part which were extracted, It consists of an actuation device interface means 8 used as an interface with the target actuation device 10, and the appliance

THIS PAGE BLANK (USPTO)

control equipment characterized by said thing [ having been recognized ] which move and/or controls the actuation device 10 of a controlled system by change of a configuration is offered. [0007] Moreover, the photography means 2 and a motion detection means 5 to detect a motion of the image photoed by said photography means, A motion partial extract means 6 to extract a part with said detected image lost motion, A window setting means 7 to set a window as the field of a predetermined motion part from said image, An image recognition means 9 to recognize change of the image which exists in said window, It consists of an actuation device interface means 8 used as an interface with the target actuation device 10, and the appliance control equipment characterized by controlling the actuation device of a controlled system by change of the image in said window is offered.

[0008] Furthermore, a photographic subject is photoed, a motion of said photoed photographic subject image is detected, a part with said detected photographic subject image lost motion is extracted, and the appliance control approach characterized by controlling the device of a controlled system according to a motion of the image of this part and/or change of a configuration is offered.

[0009] Furthermore, the distance of the image which is moving is judged from two or more images photoed by said two or more photography means 2a and 2b, and the appliance control equipment characterized by using the image which exists in a predetermined range is offered.

[0010] Furthermore, the appliance control equipment characterized by performing the judgment of said distance which is moving in said two or more images using an epipolar constraint again is offered.

[0011] By such configuration, even if a user does not move in front of an input unit or does not use the remote control equipment which transmits a control signal with infrared radiation etc., he becomes possible [ controlling a device by bodily / some /, for example, a motion of a hand, ].

[0012] Moreover, the distance of the image which is moving using the epipolar constraint using the image reflected in two or more cameras is judged, and in detection of a motion of the hand as a photographic subject, image recognition with a high precision can be performed by controlling using the image which exists in a predetermined range, preventing malfunction of the image recognition by the pattern of a background etc.

[0013]

[Embodiment of the Invention] Hereafter, the concept is explained about the appliance control equipment and the approach concerning this invention. In this invention, motion detection of the image photoed with the camera is performed first, and a window is applied to a part with a motion. Next, the image in a window is recognized and a device is operated by the recognition result.

[0014] In this case, the camera is installed as a control equipment and has always caught the motion of a user. He is trying to recognize a motion of a user's hand as a motion of a user. In addition, although operating by the 29.97Hz frame rate is [ that what is necessary is just to use a CCD camera ] desirable like the image of NTSC as for a camera, about 10Hz is sufficient.

[0015] The AD translation of the photoed image data is carried out, and the digitized frame image data incorporates two continuous frames to a frame memory. And it detects by moving to the image data of these two frames. As shown in drawing 1, supposing the flower a which is an object image was moving toward the upper right in two frames which continue by time amount t, motion detection is of operation [ which detects where block b' with the flower a of a right-hand side current image existed in the image of one frame ago ], and was in that of one frame ago in this case in the location of b.

[0016] or [ that the block of each resembles which block most by the image in front of one in fact by dividing a current image into the block of an about / 16 pixel x16 pixel ] -- difference -- taking -- pattern matching -- carrying out -- the difference -- it asks for total of data, and the smallest location (address) of the square sum. Since the number of a horizontal pixel and perpendicular direction pixels shows the address in many cases, this may be called a motion vector.

[0017] When the retrieval area of pattern matching was called search range, and horizontal resolution is 720 pixels in a 29.97Hz frame rate and it separates one frame, it is performed by

THIS PAGE BLANK (JSPTO)



about  $\approx 15$  pixels. This technique is a technique already carried out with image data compression techniques, such as MPEG.

[0018] For example, it is assumed that Hand c was shaken at right and left in the area currently photoed with the camera as shown in drawing 2. Then, a motion vector shows an effective size of population only to the block of the part which shook Hand c. That is, a motion vector shows an effective size of population, and since only the effective block d shown in drawing 3 is carrying out abbreviation quiescence of the other block, it does not have the value of an effective-travel vector. To the data of the horizontal direction (X pixel) of an actual vector, and a perpendicular direction (Y pixels), it considers as the magnitude of a vector, and the square root of each square sum is confirmed, when the magnitude is eight or more. Then, a rectangular window is set up to the image equivalent to the part which moved.

[0019] The block with which this window has an effective-travel vector value as shown in drawing 4 sets the upper and lower sides and right-and-left **\*\*\*\*\* rare \*\*** area as a rectangle. And it detects that a motion vector has eight or more magnitude, and the image in a window is captured. The condition of having upraised one finger and having turned upwards by this after shaking Hand c predetermined time is mostly incorporated like drawing 5 in the image of only the part above a wrist.

[0020] Next, the image captured as shown in drawing 6 is recognized. Here, it recognizes how many fingers of Hand c stand, whether the direction has turned to the top, or the bottom is turned to. The image on the right-hand side of drawing 6 is an image to which the captured image of Hand c was expanded. In this expansion image, level change of the luminance signal of the horizontal line A which has about 1/in the location of 4 from a top, and the horizontal line B which has about 1/in the location of 4 from the bottom is detected.

[0021] Drawing 7 shows the example of detection of level change of the luminance signal in Line A and Line B. This is an example when the intensity level of a background is larger than the intensity level of the part of Hand c (bright). A of drawing 7 is considered that one finger of Hand c is taken out in order that change of brightness may break out frequently, and since change of brightness hardly breaks out among window width, if B of drawing 7 is equivalent to the palm, it can be distinguished. It recognizes whether Hand c has turned to the top by this, or the bottom is turned to. Moreover, it is how many the part equivalent to the finger of A exists in the part equivalent to the palm of B, and it is possible to recognize how many fingers stand in the image.

[0022] Since the direction (upper and lower sides) which the number of the fingers which stand, and a finger point out by this method can be recognized, ten kinds of control can be performed in all. For example, after shaking Hand c several times, [ among these, ] (1) one finger It turns up... the switch-(2) 1 finger of TV It turns down... The switch-(3) 2 finger of TV is turned up... (4) 2 finger which carries out the increment (+) of the channel of TV is turned down... It is possible to execute a command, such as to carry out the decrement (-) of the channel to TV.

[0023] Next, the approach for improving the precision of such image recognition is explained. When Hand c is shaken, in order to distinguish the image in the background, and the image of Hand c, distance with the shaken hand c which are a camera and a photographic subject is recognized, and only the image equivalent to the distance is detected. For that purpose, two cameras can be used and the distance of an image can be detected by using the parallax of the image reflected in the two cameras. The principle is explained below.

[0024] The three-dimension picture input device using a stereo image is indicated by "graphics and the vision" (Ohm-Sha \*\*:P28.). As this shows drawing 8, the method of asking for correspondence of each pixel of the image of the right and left which kept the fixed distance L between opticals axis, and have been arranged, and finding the distance from the parallax to the point of measurement corresponding to the image is adopted. What is necessary is just to perform corresponding-points retrieval on the same scanning line which is an epipolar line, since it arranges here so that the optical axis of two cameras may usually be included on the same flat surface in the three-dimension picture input device using a stereo image (epipolar constraint). That is, when distance between the opticals axis of a camera was set to L, the focal distance was set to f and the point P on an object (x y, z) is projected on the points Pl (Xl, Yl) and Pr (Xr, Yr) on a screen on either side, respectively, distance z is  $z = Lxf / (Xl - Xr)$ .

THIS PAGE BLANK (USP:0)

It is come out and expressed. Here, when  $XI-Xr$  expresses parallax and a left image is used as a criteria image, the parallax vector in Point  $PI(XI, YI)$  is expressed as  $****(XI-Xr, YI-Yr)$ .

[0025] Since it becomes possible to detect the distance for every pixel of an image according to this approach For example, as shown in drawing 9 , one camera 2 is installed in TV1 for monitors. When it is going to recognize Hand c by the approach mentioned above, furniture 3 and a curtain 4 exist in the background. Even when detection of the finger of Hand c is performed good neither by furniture 3 nor the pattern of a curtain 4 As shown in drawing 10 , when two cameras, camera 2a and camera 2b, are installed in TV1 for monitors, The parallax vector of the image photoed by two camera 2a and 2b is calculated using the above-mentioned epipolar constraint, and it becomes possible to extract only the image which exists in the distance of the part which detected and mentioned above camera 2a of Hand c, and the distance from 2b, and which moved and carried out distance detection with detection.

[0026] That is, if the part which fulfills two conditions of being the field which is a motion field and exists in the same distance as the part which moved, from the image with which Hand c and furniture 3 which are a photographic subject for transmitting a command as shown in drawing 11 , and a curtain 4 existed is extracted, as shown in drawing 12 , only the part of Hand c can be extracted. And it becomes possible to perform vertical recognition of Hand c, and number recognition of a finger with a sufficient precision more by embedding black, white, etc. in the field where the distance of a background is large.

[0027] The appliance control equipment of the 1st example which starts this invention using a technique which was mentioned above is explained using drawing 13 . First, the image photoed with the camera 2 is transmitted to the motion detector 5. The image transmitted to the motion detector 5 is divided into a predetermined block, it moves for every frame, and detection is made. The motion vector information for every block by which motion detection was carried out is transmitted to the motion partial extractor 6. Next, in the command transmitting photographic subject window setting section 7, the window area of a rectangle including the block detected as a block with a motion is set up, and the image which exists in the window area is transmitted to the image recognition machine 9. With the image recognition vessel 9, the upper and lower sides of the direction which the number of the finger of the photographic subject (hand c) for transmitting the command mentioned above and a finger point out are recognized. The recognized information is sent to the actuation device interface 8, and is changed into the control signal over the actuation devices 10 set up beforehand, such as ON/OFF of a switch. And the actuation device 10 is controlled by this control signal.

[0028] Next, as shown in drawing 10 , two cameras are explained using drawing 14 about the appliance control equipment of the 2nd example which starts this invention corresponding to the case where it installs at the predetermined spacing. The image photoed by two camera 2a and 2b is transmitted to the epipolar restricted parallax vector extractor 11, respectively. In the epipolar restricted parallax vector extractor 11, the parallax vector of two images is calculated using an epipolar constraint, and it transmits to the distance calculation machine 12. With the distance calculation vessel 12, the distance from each camera side of the photographic subject (hand c) which transmits a command is detected from the inputted parallax vector, and the distance information for every pixel is transmitted to the command transmitting photographic subject window setter 7. The image photoed by camera 2b on the other hand is transmitted to the motion detector 5. The image transmitted to the motion detector 5 is divided into a predetermined block, for every frame, moves and is detected. And it is transmitted to the motion partial extractor 6, and the motion vector information for every block by which motion detection was carried out extracts the block which calculates the absolute value of a motion vector here and has eight or more values.

[0029] The address of the extracted block is transmitted to the command transmitting photographic subject window setting section 7. In the command transmitting photographic subject window setting section 7 Based on the distance information for every [ which moved and was inputted from the address information and the distance calculation machine 12 of a block ] pixel inputted from the motion partial extractor 6 Extract an image part equal to the distance of the part which is moving, and the window area of a rectangle including the block detected as a

THIS PAGE BLANK (USPTO)

block with a motion is set up. Background images other than the photographic subject which transmits a command are set as white (pixel level 255), and the image which exists in the window is transmitted to the image recognition machine 9. In the image recognition section 9, the upper and lower sides of the direction which the number of the finger the hand c which is the photographic subject which transmits the command mentioned above stands, and a finger point out are recognized. It is transmitted to the actuation device interface 8, and the number information on a finger and the information on up-and-down which have been recognized are changed into the control signal corresponding to appliance control set up beforehand, such as ON/OFF of a switch. And a control signal is transmitted to the actuation device 10, and the actuation device 10 is controlled.

[0030] Next, two cameras are used, and when controlling two or more devices, the appliance control equipment of the 3rd example which starts [\*\*\*\*\*] this invention is explained using drawing 15.

[0031] The image photoed by two camera 2a and 2b is transmitted to the epipolar restricted parallax vector extractor 11, respectively. In the epipolar restricted parallax vector extractor 11, the parallax vector of two images photoed by two camera 2a and 2b is calculated using an epipolar constraint, and it transmits to the distance calculation machine 12. With the distance calculation vessel 12, the distance from each camera side of the hand c which is a photographic subject for transmitting a command is detected from the inputted parallax vector, and the distance information for every pixel is transmitted to the command transmitting photographic subject window setter 7. On the other hand, the image photoed by camera 2b is transmitted to the motion detector 5. The image transmitted to the motion detector 5 is divided into a predetermined block, for every frame, moves and is detected. It is transmitted to the motion partial extractor 6, and the motion vector information for every block by which motion detection was carried out calculates the absolute value of a motion vector, and he is trying to extract the block with eight or more values in the motion partial extractor 6. Thus, the address of the extracted block is transmitted to the command transmitting photographic subject window setter 7.

[0032] In the command transmitting photographic subject window setter 7 It carries out based on the address information of the inputted motion block, and the distance information for every pixel. Extract an image part equal to the distance of the part which is moving, and the window area of a rectangle including the block detected as a block with a motion is set up. Background images other than a command transmitting photographic subject are set as white (pixel level 255), and the image which exists in the window is transmitted to the image recognition machine 9. With the image recognition vessel 9, the upper and lower sides of the direction which the number of the finger the hand c which is a photographic subject for transmitting the command mentioned above stands, and a finger point out are recognized. And it is transmitted to the actuation device judging machine 13, and such information specifies an actuation device.

[0033] For example, after shaking a hand several times, (1) 1 finger is turned up in the location... The switch of the actuation device A is turned ON.

(2) Turn one finger down... The switch of the actuation device A is turned OFF.

(3) Turn two fingers up... The switch of the actuation device B is turned ON.

(4) Turn two fingers down... The switch of the actuation device B is turned OFF.

(5) Turn three fingers up... The switch of the actuation device C is turned ON.

(6) Turn three fingers down... The switch of the actuation device C is turned OFF.

(7) Turn four fingers up... The switch of the actuation device D is turned ON.

(8) Turn four fingers down... The switch of the actuation device D is turned OFF.

supposing it is assigning the command to say, specify an actuation device at the number of the finger which stands, and by the specified actuation device The information on the upper and lower sides (the direction of Hand c) of a finger is transmitted to actuation device interfaces [linked to the target device / 8a-8d] either. It is changed into the control signal corresponding to appliance control set up beforehand, such as ON/OFF of a switch, and the signal is transmitted to control equipments [10a-10d] either, and control of each device is performed. Here, as an actuation device, if TV, VTR, an air-conditioner, certification, etc. can use remote

THIS PAGE BLANK (USPTO)

control, it can be adapted.

[0034] In addition, although this example explained as a photographic subject for transmitting a command using the hand which are some human being's bodies, a face, a motion of an eye, etc. may be used according to the mystery of the technique of image recognition. Moreover, it is not some human being's bodies and predetermined printed matter, a predetermined bar code, etc. of a pattern may be used.

[0035]

[Effect of the Invention] As mentioned above, according to the appliance control equipment and the approach concerning this invention, as explained in full detail, even if a user does not move in front of an input unit or it does not use remote control equipment separately, it becomes possible to control a device. Moreover, it can prevent the precision of image recognition improving and controlling a device by using two or more cameras accidentally.

---

[Translation done.]

THIS PAGE BLANK (USP 10)



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-5975

(P2001-5975A)

(43) 公開日 平成13年1月12日 (2001.1.12)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テームト* (参考)
G 0 6 T 7/20		G 0 6 F 15/70	4 1 0 5 C 0 5 6
G 0 6 F 3/00	6 8 0	3/00	6 8 0 C 5 K 0 4 8
H 0 4 N 5/00		H 0 4 N 5/00	A 5 L 0 9 6
H 0 4 Q 9/00	3 3 1	H 0 4 Q 9/00	3 3 1 A

審査請求 未請求 請求項の数 5 O L (全 11 頁)

(21) 出願番号 特願平11-173747

(22) 出願日 平成11年6月21日 (1999.6.21)

(71) 出願人 000004329

日本ビクター株式会社

神奈川県横浜市神奈川区守屋町3丁目12番  
地

(72) 発明者 菅原 隆幸

神奈川県横浜市神奈川区守屋町3丁目12番  
地 日本ビクター株式会社内

Fターム (参考) 5C056 AA04 BA01 BA08 DA08

5K048 BA03 DB02 EB02

5L096 BA08 CA04 CA05 EA35 FA64

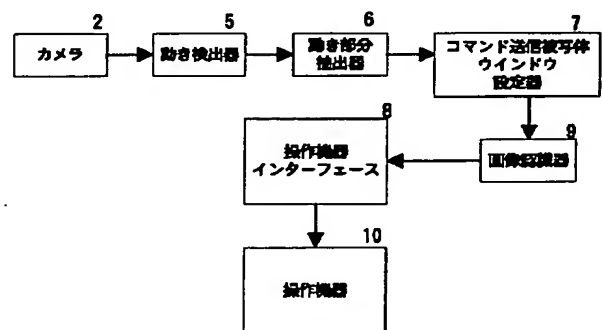
FA66 FA81 GA08 GA17 HA04

(54) 【発明の名称】 機器制御装置及び方法

(57) 【要約】

【課題】 ユーザーが入力装置やリモコン装置を使用することなく機器の制御を可能とする。

【解決手段】 撮影手段2と、これによって撮影された画像の動きを検出する動き検出手段5と、検出された画像から動きのある部分を抽出する動き部分抽出手段6と、抽出された部分の画像の動き及び／又は形状を認識する画像認識手段9と、対象の操作機器10とのインターフェースとなる操作機器インターフェース手段8とからなり、認識された動き及び／又は形状の変化によって制御対象の操作機器を制御する。



## 【特許請求の範囲】

## 【請求項1】撮影手段と、

前記撮影手段によって撮影された画像の動きを検出する動き検出手段と、

前記検出された画像から動きのある部分を抽出する動き部分抽出手段と、

前記抽出された部分の画像の動き及び／又は形状を認識する画像認識手段と、

対象の操作機器とのインターフェースとなる操作機器インターフェース手段とからなり、前記認識された動き及び／又は形状の変化によって制御対象の操作機器を制御することを特徴とする機器制御装置。

## 【請求項2】撮影手段と、

前記撮影手段によって撮影された画像の動きを検出する動き検出手段と、

前記検出された画像から動きのある部分を抽出する動き部分抽出手段と、

前記画像から所定の動き部分の領域にウィンドウを設定するウィンドウ設定手段と、

前記ウィンドウ内に存在する画像の変化を認識する画像認識手段と、

対象の操作機器とのインターフェースとなる操作機器インターフェース手段とからなり、前記ウィンドウ内の画像の変化によって制御対象の操作機器を制御することを特徴とする機器制御装置。

## 【請求項3】被写体を撮影し、

前記撮影された被写体画像の動きを検出し、

前記検出された被写体画像から動きのある部分を抽出して、この部分の画像の動き及び／又は形状の変化に応じて制御対象の機器を制御することを特徴とする機器制御方法。

【請求項4】複数の前記撮影手段によって撮影された複数の画像から、動いている画像の距離を判定し、所定の距離範囲に存在する画像を用いたことを特徴とする請求項1又は請求項2に記載の機器制御装置。

【請求項5】前記動いている距離の判定は、前記複数の画像においてエッジや拘束条件を使用して行うことを特徴とする請求項4に記載の機器制御装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、機器を遠隔制御するための装置及び方法に関し、特にカメラで検出した画像情報、動き情報、又は双方の情報をを用いて機器の制御を行う機器制御装置及び方法に関する。

## 【0002】

【従来の技術】従来、テレビやビデオ等に用いられている赤外線を利用したリモコンがある。また、CCDカメラ等を使用して、画像の動きを認識し、動きが検出された場合に、カメラに接続しているVTRの録画を開始させる監視システム等の装置が存在している。

【0003】特開平5-307793号公報に開示された遠隔制御装置は、図16に示すように、使用者とのインターフェースとなり、表示内容を変更可能な表示・入力部100と、マイコン（演算処理部）200と、VTR等の制御対象部400と、メモリ等の記憶手段300とで構成されている。そして、表示・入力部100によって指示された命令を、演算処理部200によって演算処理し、命令を解釈した後実行し、メモリ300に新しい表示・入力部100のための情報を保存して、その情報により表示・入力部100を使用状態によって可変できるようにしている。これは、制御対象部400を使用している間に、使用者の使用状態（癖や特徴や傾向など）を学習し、より使いやすいインターフェースに自動的に設定する目的で発明されたものである。

## 【0004】

【発明が解決しようとする課題】しかしながら、上記した従来の機器制御装置では、入力装置としてユーザーとのインターフェース用のボタンやタッチパネルが必要であった。このような入力装置では、ユーザーが入力装置の前まで移動して操作する必要があった。また、ユーザーが入力装置の前まで移動せずに操作するために赤外線などの制御信号を送信するリモートコントロール（遠隔制御）装置が必要であったが、リモートコントロール装置がユーザーの近くに無かったり、一時的に紛失することで操作上の不便を感じることも多々あった。

【0005】更に、監視システムにおいて、カメラによって撮影した画像の動きを検出してVTRの録画を開始するように制御する方法があるが、これはあくまでも画像の動きの有無を検出して簡単な制御を行うものであり、ユーザーインターフェースとしての役割を果たしているとは言い難いものであった。

## 【0006】

【課題を解決するための手段】上述した課題を解決するために、撮影手段2と、前記撮影手段2によって撮影された画像の動きを検出する動き検出手段5と、前記検出された画像から動きのある部分を抽出する動き部分抽出手段6と、前記抽出された部分の画像の動き及び／又は形状を認識する画像認識手段9と、対象の操作機器10とのインターフェースとなる操作機器インターフェース手段8とからなり、前記認識された動き及び／又は形状の変化によって制御対象の操作機器10を制御することを特徴とする機器制御装置を提供する。

【0007】また、撮影手段2と、前記撮影手段2によって撮影された画像の動きを検出する動き検出手段5と、前記検出された画像から動きのある部分を抽出する動き部分抽出手段6と、前記画像から所定の動き部分の領域にウィンドウを設定するウィンドウ設定手段7と、前記ウィンドウ内に存在する画像の変化を認識する画像認識手段9と、対象の操作機器10とのインターフェースとなる操作機器インターフェース手段8とからなり、前記

ウインドウ内の画像の変化によって制御対象の操作機器を制御することを特徴とする機器制御装置を提供する。

【0008】更に、被写体を撮影し、前記撮影された被写体画像の動きを検出し、前記検出された被写体画像から動きのある部分を抽出して、この部分の画像の動き及び／又は形状の変化に応じて制御対象の機器を制御することを特徴とする機器制御方法を提供する。

【0009】また更に、複数の前記撮影手段2a、2bによって撮影された複数の画像から、動いている画像の距離を判定し、所定の距離範囲に存在する画像を用いたことを特徴とする機器制御装置を提供する。

【0010】更にまた、前記動いている距離の判定は、前記複数の画像においてエピソード拘束条件を使用して行うことを特徴とする機器制御装置を提供する。

【0011】このような構成によって、ユーザーは入力装置の前まで移動したり、赤外線などで制御信号を送信するリモートコントロール装置を使用しなくても、体の一部、例えば手の動きによって機器を制御することが可能となる。

【0012】また、複数のカメラに映っている画像を用いてエピソード拘束条件を使用して動いている画像の距離を判定し、所定の距離範囲に存在する画像を用いて制御を行うことで、被写体としての手の動きの検出において、背景の模様等による画像認識の誤動作を防止しながら精度の高い画像認識を行うことができる。

【0013】

【発明の実施の形態】以下、本発明に係る機器制御装置及び方法について、その概念を説明する。本発明では、まず、カメラによって撮影した画像の動き検出を行い、動きのある部分にウインドウをかける。次に、ウインドウ内の画像を認識し、その認識結果によって機器を操作する。

【0014】この場合、カメラは制御機器として設置されており、常にユーザーの動きを捉えている。ユーザーの動きとしてはユーザーの手の動きを認識するようにしている。なお、カメラはCCDカメラを使用すれば良く、NTSCの画像のように、29.97Hzのフレームレートで作動しているのが好ましいが、10Hz程度でも構わない。

【0015】撮影された画像データはAD変換され、デジタル化されたフレーム画像データは連続する2フレームをフレームメモリに取り込む。そして、この2つのフレームの画像データに対して動き検出を行う。動き検出は図1に示すように、時間tで連続する2つのフレームの中で対象画像である花aが右上に向かって動いていたとすると、右側のカレント画像の花aのあるブロックb'が1フレーム前の画像においてどこに存在していたかを検出する動作のことであり、この場合は1フレーム前にはbの位置にあった。

【0016】実際には、カレント画像を16画素×16

画素程度のブロックに分けて、その一つ一つのブロックが一つ前の画像でどのブロックに一番似ているかを、差分を取ってパターンマッチングをし、その差分データの総和や、2乗和の最も小さい場所（アドレス）を求める。アドレスは水平方向画素と垂直方向画素の数で示すことが多いので、これを動きベクトルと呼ぶこともある。

【0017】パターンマッチングの検索エリアは、サーチレンジといって、29.97Hzのフレームレートで水平方向解像度が720画素の場合、1フレーム離れたとき、±15画素程度で行う。この技術はMPEGなどの画像データ圧縮技術にて既に実施されている技術である。

【0018】例えば、図2に示すようにカメラによって撮影されているエリア内で手cを左右に振ったと仮定する。すると、手cを振った部分のブロックにのみ動きベクトルは有効な大きさを示す。即ち、図3に示す有効ブロックdのみ動きベクトルは有効な大きさを示し、それ以外のブロックは略静止しているため有効な動きベクトルの値を持たない。実際のベクトルの水平方向（X画素）、垂直方向（Y画素）のデータに対して、それぞれの2乗和の平方根を、そのベクトルの大きさとし、その大きさが8以上の場合に有効であるとする。そこで、動いた部分に相当する画像に対して長方形のウインドウを設定する。

【0019】このウインドウは図4に示すように、有効な動きベクトル値を有するブロックが上下、左右全て含まれるエリアを長方形に設定する。そして、動きベクトルが8以上の大きさを持つことを検出し、ウインドウ内の画像を取り込む。これによって、手cを所定の時間振った後、指を1本立てて上に向けた状態の、ほぼ手首より上の部分だけの画像が図5のように取り込まれる。

【0020】次に、図6に示すように取り込んだ画像の認識を行う。ここでは、手cの指が何本立っているか、また、その方向が上を向いているか、下を向いているかを認識する。図6の右側の画像は取り込んだ手cの画像を拡大した画像である。この拡大画像において、上から約1/4の位置にある水平方向の線Aと、下から約1/4の位置にある水平方向の線Bとの輝度信号のレベル変化を検出する。

【0021】図7は線Aと線Bにおける輝度信号のレベル変化の検出例を示している。これは、背景の輝度レベルの方が手cの部分の輝度レベルよりも大きい（明るい）場合の例である。図7のAは輝度の変化が頻繁に起きるため、手cの指が1本出されていると考えられ、図7のBは輝度の変化がウインドウ幅のうちほとんど起きないので手のひらに相当していると判別できる。これによって手cが上を向いているか、下を向いているかを認識する。また、Bの手のひらに相当する部分の中にAの指に相当する部分がいくつ存在するかで、画像内に何本

の指が立っているかを認識することが可能である。

【0022】この方式で、立っている指の数、指が指す方向（上下）を認識できるので、全部で10種類の制御を実行することができる。例えば、このうち、手cを数回振ったあとで、

- (1) 1本の指を上にする・・・TVのスイッチON
- (2) 1本の指を下にする・・・TVのスイッチOFF
- (3) 2本の指を上にする・・・TVのチャンネルをインクリメント(+)する
- (4) 2本の指を下にする・・・TVにチャンネルをデクリメント(-)する

などのコマンドを実行することが可能である。

【0023】次に、このような画像認識の精度を向上するための方法を説明する。手cを振ったときにその背景にある画像と手cの画像とを区別するために、カメラと被写体である振られた手cとの距離を認識し、その距離に相当する画像のみを検出する。そのためにはカメラを2台使用し、その2台のカメラに映る画像の視差を用いることで、画像の距離を検出することができる。その原理を以下に説明する。

【0024】ステレオ画像を用いた3次元画像入力装置は、例えば、「グラフィックスとビジョン」（オーム社刊：P28）に記載されている。これは図8に示すように、光軸間に一定の距離Lを置いて配置された左右の画像の各画素の対応を求め、その視差からその画像に対応する測定点までの距離を求める方法が採用されている。ここで、ステレオ画像を用いた3次元画像入力装置では、通常、2台のカメラの光軸が同一平面上に含まれるように配置するので、対応点探索はエピポーラ線である同じ走査線上で行えば良い（エピポーラ拘束条件）。すなわち、カメラの光軸間の距離をL、焦点距離をfとすると、対象物上の点P（x、y、z）が左右の画面上の点P<sub>l</sub>（X<sub>l</sub>、Y<sub>l</sub>）、P<sub>r</sub>（X<sub>r</sub>、Y<sub>r</sub>）にそれぞれ投影されたとき、距離zは

$$z = L \times f / (X_l - X_r)$$

で表される。ここで、X<sub>l</sub> - X<sub>r</sub>は視差を表し、向かって左の画像を基準画像とした場合、点P<sub>l</sub>（X<sub>l</sub>、Y<sub>l</sub>）における視差ベクトルはV<sub>p</sub>（X<sub>l</sub> - X<sub>r</sub>、Y<sub>l</sub> - Y<sub>r</sub>）と表される。

【0025】この方法によれば、画像の画素毎の距離を検出することが可能になるので、例えば図9に示すように、モニター用のTV1にカメラ2が1台設置されており、前述した方法で手cを認識しようとした場合、その背景に家具3やカーテン4が存在していて、家具3やカーテン4の模様などで手cの指の検出が良好に行われないうちでも、図10に示すようにモニター用のTV1にカメラ2a及びカメラ2bの2台のカメラが設置されている場合、2台のカメラ2a、2bによって撮影された画像の視差ベクトルを上記エピポーラ拘束条件を使用して演算し、手cのカメラ2a、2bからの距離を検出

し、前述した動き検出と共に距離検出した部分の距離に存在する画像のみを抽出することが可能となる。

【0026】即ち、図11に示すように、コマンドを送信するための被写体である手cと家具3とカーテン4が存在していた画像から、動き領域であって、動いた部分と同じ距離に存在する領域であるという2つの条件を満たしている部分を抽出すると、図12に示すように手cの部分のみが抽出できる。そして、背景の距離が大きい領域には例えば、黒や白などを埋め込むことによって、より手cの上下認識や指の本数認識を精度良く行うことが可能となる。

【0027】上述したような技術を用いて本発明に係る第1実施例の機器制御装置について図13を使用して説明する。まず、カメラ2によって撮影された画像は動き検出器5に送信される。動き検出器5に送信された画像は所定のブロックに分割されて、フレーム毎に動き検出がなされる。動き検出されたブロック毎の動きベクトル情報は、動き部分抽出器6に送信される。次に、コマンド送信被写体ウインドウ設定部7では、動きのあるブロックとして検出されたブロックを含む長方形のウインドウ領域を設定し、そのウインドウ領域内に存在する画像を画像認識器9へ送信する。画像認識器9では前述したコマンドを送信するための被写体（手c）の指の本数及び指の指す方向の上下を認識する。認識した情報は操作機器インターフェース8へ送られ、スイッチのON/OFFなどの予め設定した操作機器10に対する制御信号に変換される。そして、この制御信号によって操作機器10が制御される。

【0028】次に、図10に示すように、2台のカメラを所定の間隔で設置した場合と対応して本発明に係る第2実施例の機器制御装置について図14を使用して説明する。2台のカメラ2a、2bによって撮影された画像はそれぞれエピポーラ拘束視差ベクトル抽出器11に送信される。エピポーラ拘束視差ベクトル抽出器11では2つの画像の視差ベクトルをエピポーラ拘束条件を使用して演算し、距離算出器12へ送信する。距離算出器12では入力された視差ベクトルからコマンドを送信する被写体（手c）の各カメラ面からの距離を検出し、画素毎の距離情報をコマンド送信被写体ウインドウ設定器7へ送信する。その一方で、カメラ2bによって撮影された画像は動き検出器5に送信される。動き検出器5に送信された画像は所定のブロックに分割されて、フレーム毎に動き検出される。そして、動き検出されたブロック毎の動きベクトル情報は動き部分抽出器6に送信され、ここで動きベクトルの絶対値を演算し8以上の値を持つブロックを抽出する。

【0029】抽出されたブロックのアドレスはコマンド送信被写体ウインドウ設定部7に送信される。コマンド送信被写体ウインドウ設定部7では、動き部分抽出器6から入力された動きブロックのアドレス情報と、距離算

出器12から入力された画素毎の距離情報を基に、動いている部分の距離と等しい画像部分を抽出し、動きのあるブロックとして検出されたブロックとを含む長方形のウィンドウ領域を設定し、コマンドを送信する被写体以外の背景画像は白(画素レベル255)に設定し、そのウィンドウ内に存在する画像を画像認識器9へ送信する。画像認識器9では前述したコマンドを送信する被写体である手cの立っている指の本数及び指の指す方向の上下を認識する。認識した指の本数情報や上下の情報は操作機器インターフェース8に送信されて、スイッチのON/OFFなどの予め設定した機器制御に対応した制御信号に変換される。そして、制御信号は操作機器10へと送信され、操作機器10が制御される。

【0030】次に、カメラを2台使用し、複数の機器を制御する場合として本発明に係る第3実施例の機器制御装置について、図15を使用して説明する。

【0031】2台のカメラ2a、2bによって撮影された画像はそれぞれエピポーラ拘束視差ベクトル抽出器11に送信される。エピポーラ拘束視差ベクトル抽出器11では、2台のカメラ2a、2bによって撮影された2つの画像の視差ベクトルをエピポーラ拘束条件を使用して演算し、距離算出器12に送信する。距離算出器12では入力された視差ベクトルからコマンドを送信するための被写体である手cの各カメラ面からの距離を検出し、画素毎の距離情報をコマンド送信被写体ウィンドウ設定器7に送信する。一方、カメラ2bによって撮影された画像は動き検出器5に送信される。動き検出器5に送信された画像は所定のブロックに分割されて、フレーム毎に動き検出される。動き検出されたブロック毎の動きベクトル情報は動き部分抽出器6に送信され、動き部分抽出器6では動きベクトルの絶対値を計算し、8以上の値を持つブロックを抽出するようにしている。このように抽出されたブロックのアドレスはコマンド送信被写体ウィンドウ設定器7に送信される。

【0032】コマンド送信被写体ウィンドウ設定器7では、入力された動きブロックのアドレス情報と、画素毎の距離情報とを基にして、動いている部分の距離と等しい画像部分を抽出し、動きのあるブロックとして検出されたブロックを含む長方形のウィンドウ領域を設定し、コマンド送信被写体以外の背景画像は白(画素レベル255)に設定し、そのウィンドウ内に存在する画像を画像認識器9へ送信する。画像認識器9では、前述したコマンドを送信するための被写体である手cの立っている指の本数、指の指す方向の上下を認識する。そして、このような情報は操作機器判定器13に送信され、操作機器を特定する。

【0033】例えば、手を数回振った後に、その位置で  
(1) 1本の指を上にする・・・操作機器AのスイッチをONにする。

(2) 1本の指を下にする・・・操作機器Aのスイッチ

をOFFにする。

(3) 2本の指を上にする・・・操作機器BのスイッチをONにする。

(4) 2本の指を下にする・・・操作機器BのスイッチをOFFにする。

(5) 3本の指を上にする・・・操作機器CのスイッチをONにする。

(6) 3本の指を下にする・・・操作機器CのスイッチをOFFにする。

(7) 4本の指を上にする・・・操作機器DのスイッチをONにする。

(8) 4本の指を下にする・・・操作機器DのスイッチをOFFにする。

というコマンドを割り当てていたとすると、立っている指の本数で操作機器を特定し、特定された操作機器では、指の上下(手cの方向)の情報が対象の機器に接続した操作機器インターフェース8a~8dのいずれかに送信され、スイッチのON/OFFなど予め設定した機器制御に対応した制御信号に変換され、その信号は制御機器10a~10dのいずれかに送信されて、各機器の制御が行われる。ここで、操作機器としては、TVやVTR、エアコン、証明など、リモコンが使用できるものであれば適応可能である。

【0034】なお、本実施例では、コマンドを送信するための被写体として、人間の体の一部である手を用いて説明したが、画像認識の技術の神秘によって、顔や目の動きなどを用いても良い。また、人間の体の一部ではなく、所定のパターンの印刷物やバーコードなどを使用しても良い。

【0035】

【発明の効果】以上、詳述したように、本発明に係る機器制御装置及び方法によれば、ユーザーが入力装置の前に移動したり、別途リモートコントロール装置を使用しなくても機器を制御することが可能となる。また、カメラを複数使用することで、画像認識の精度が向上し、機器を誤って制御することを防止することができる。

【図面の簡単な説明】

【図1】動き検出を説明するための図である。

【図2】動き部分の抽出を説明するための図である。

【図3】動き部分の抽出を説明するための図である。

【図4】コマンド送信被写体のウィンドウの設定を説明するための図である。

【図5】コマンド送信被写体のウィンドウの設定を説明するための図である。

【図6】コマンド送信被写体の画像認識方法を説明するための図である。

【図7】コマンド送信被写体の画像認識方法を説明するための図である。

【図8】3次元入力とエピポーラ拘束条件を説明するための図である。

【図9】カメラが2台ある場合のカメラ設定位置の説明図である。

【図10】カメラが2台ある場合のカメラ設定位置の説明図である。

【図11】カメラが2台ある場合のウィンドウ設定を説明するための図である。

【図12】カメラが2台ある場合のウィンドウ設定を説明するための図である。

【図13】本発明に係る第1実施例の機器制御装置の一実施例を示すブロック図である。

【図14】本発明に係る第2実施例の機器制御装置の一実施例を示すブロック図である。

【図15】本発明に係る第3実施例の機器制御装置の一実施例を示すブロック図である。

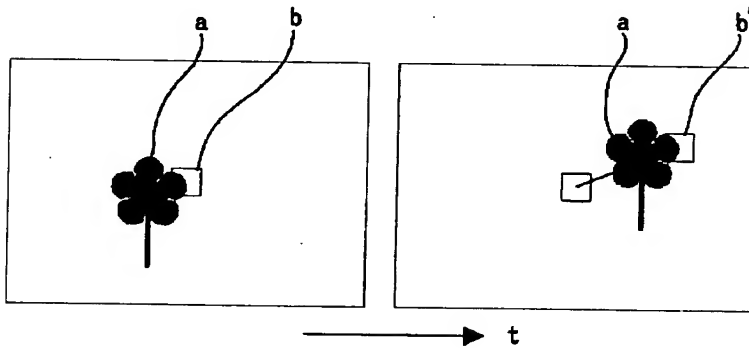
【図16】従来の機器制御装置を示すブロック図であ

る。

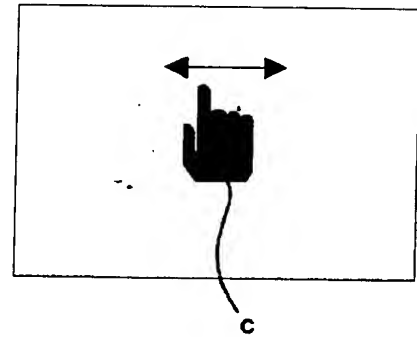
【符号の説明】

- 1 TV
- 2, 2 a, 2 b カメラ
- 3 家具
- 4 カーテン
- 5 動き検出器
- 6 動き部分抽出器
- 7 コマンド送信被写体ウィンドウ設定部
- 8 操作機器インターフェース
- 9 画像認識器
- 10 操作機器
- 11 エピポーラ拘束視差ベクトル抽出器
- 12 距離算出器
- 13 操作機器判別器

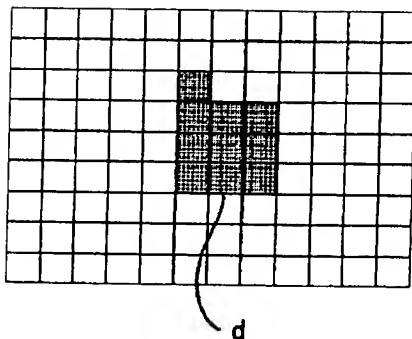
【図1】



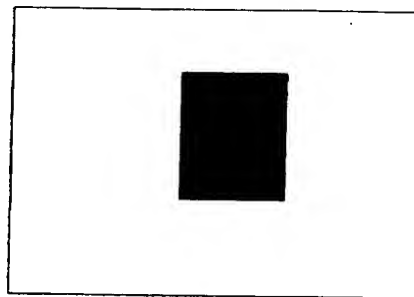
【図2】



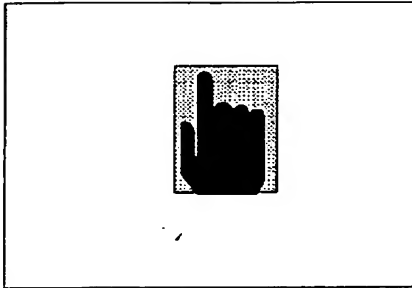
【図3】



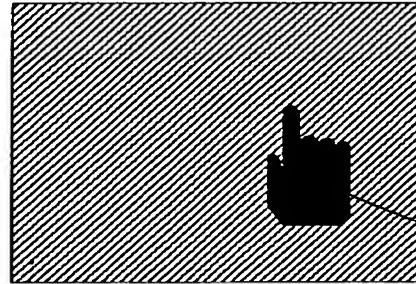
【図4】



【図5】

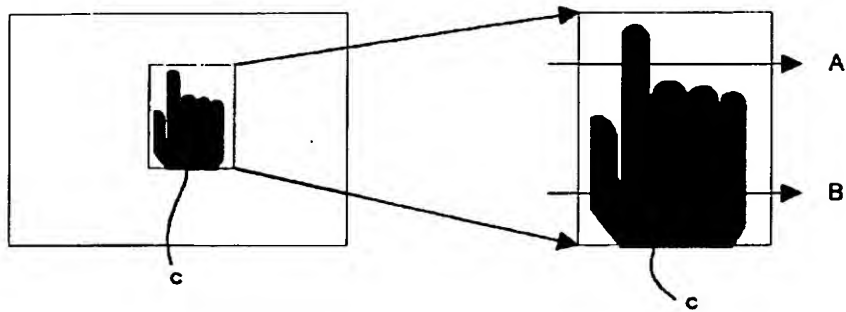


【図12】

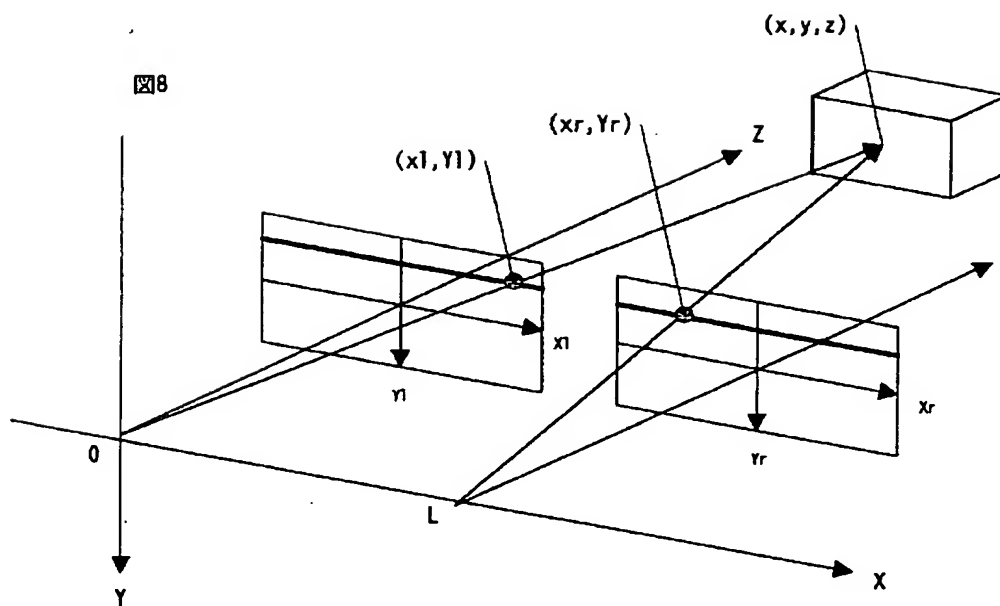


コマンド送信被写体 C

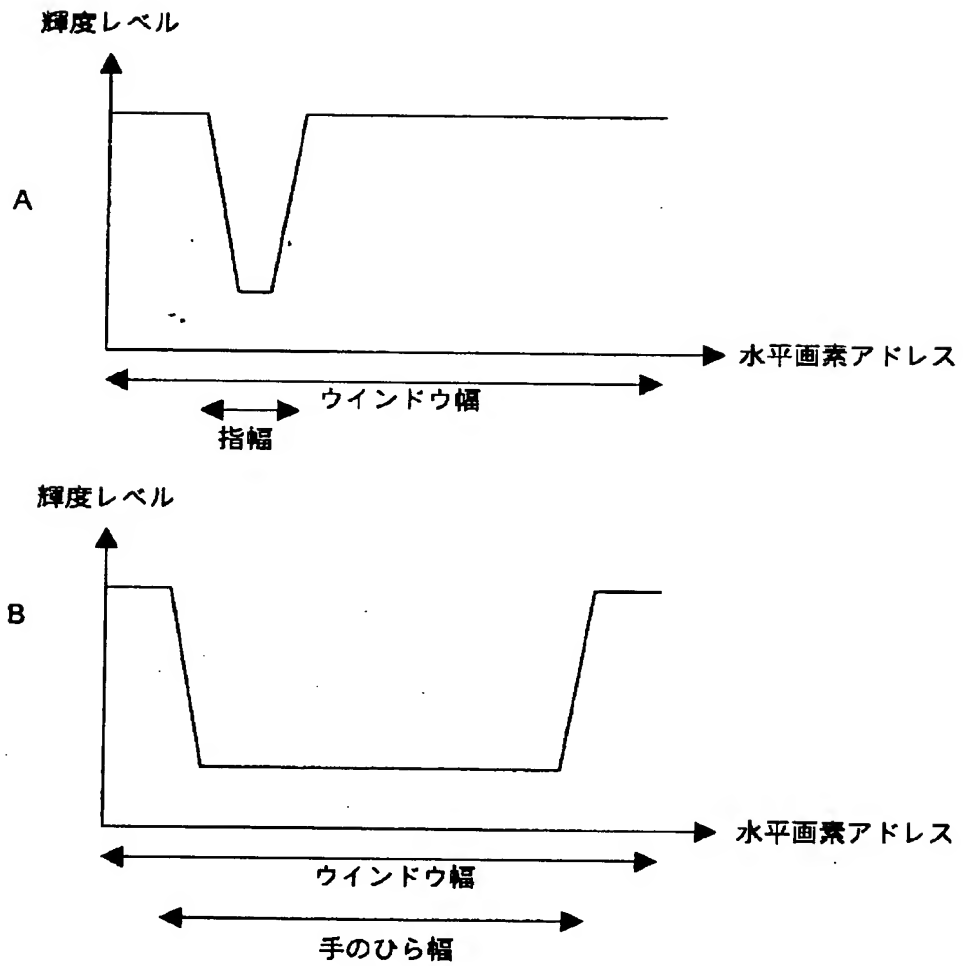
【図6】



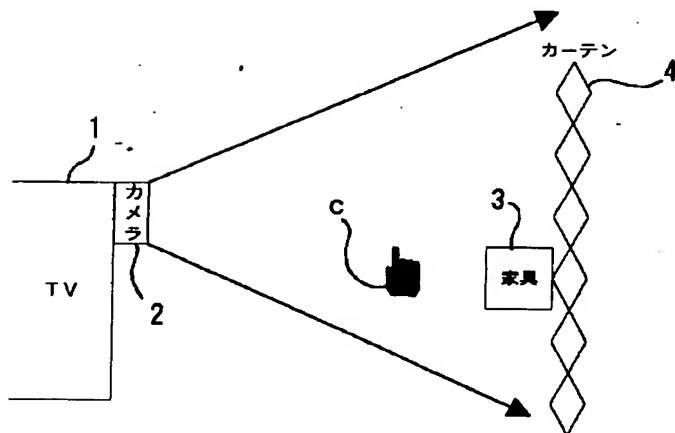
【図8】



【図7】

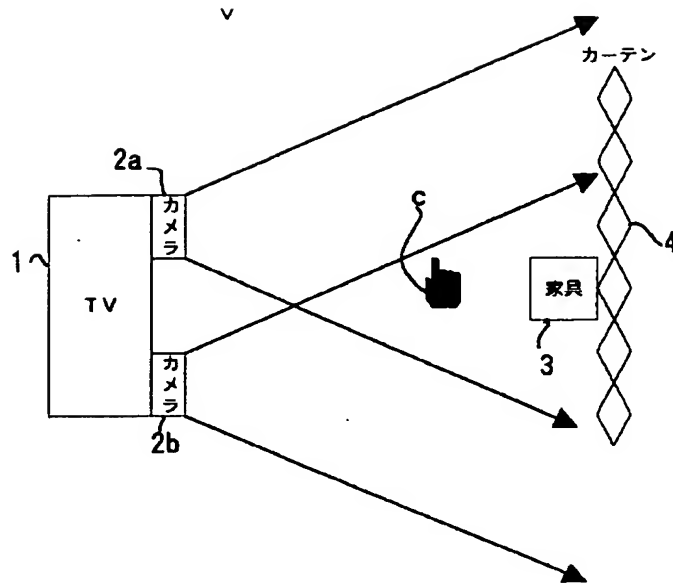


【図9】

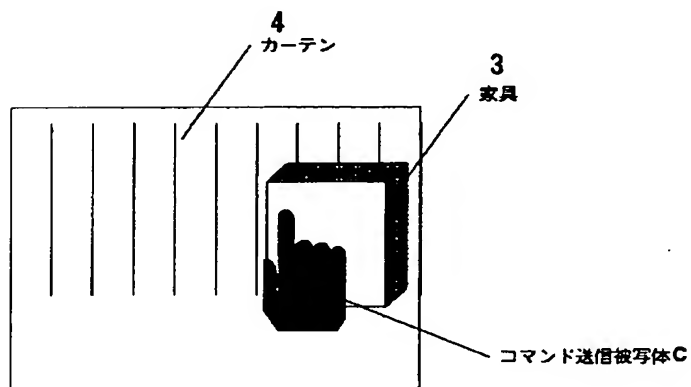




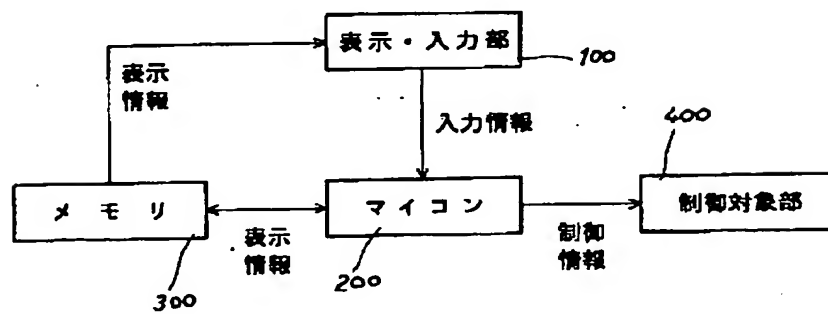
【図10】



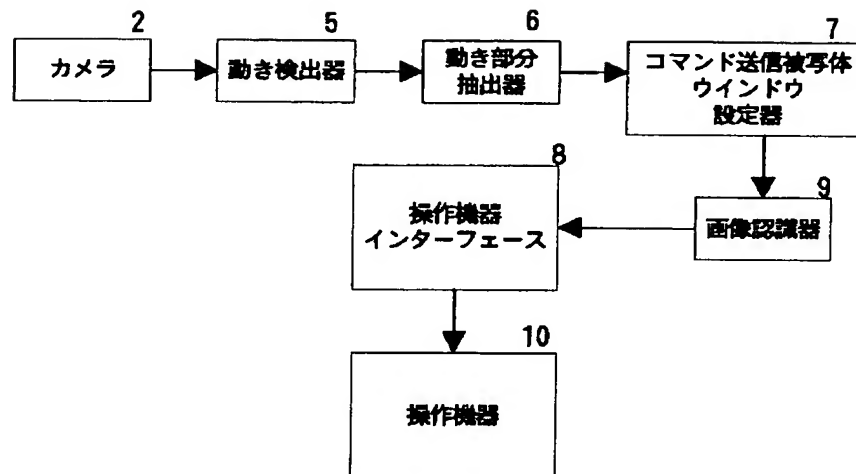
【図11】



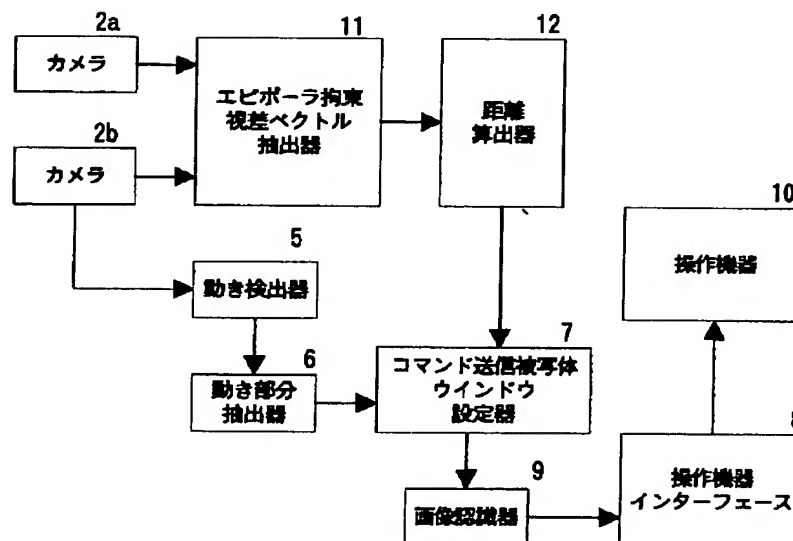
【図16】



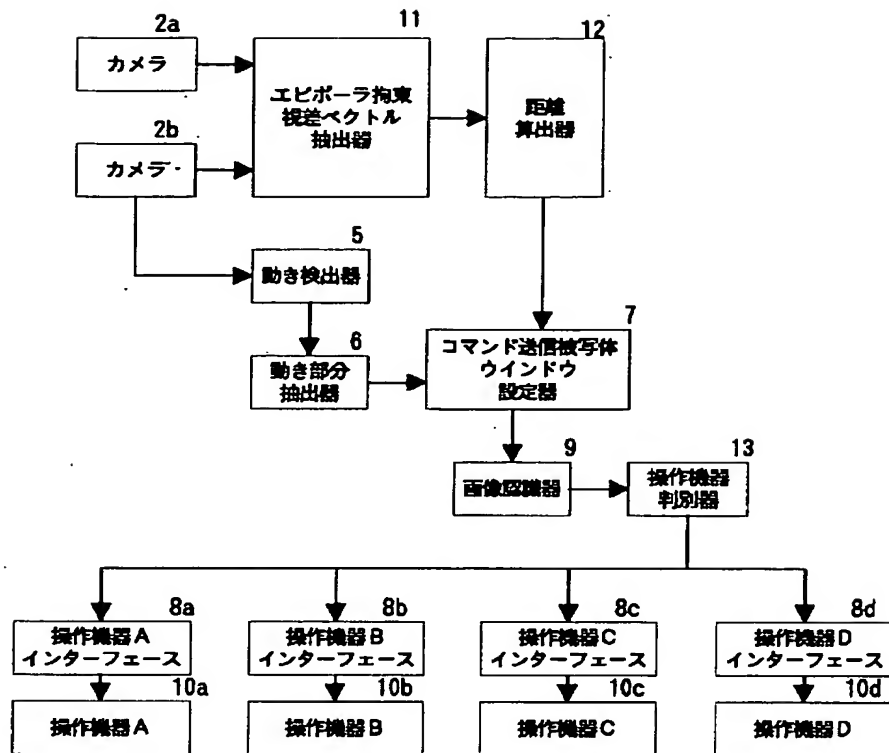
【図13】



【図14】



【図15】



THIS PAGE BLANK (USPTO)